Supplement to "Multiscale Clustering of Nonparametric Regression Curves"

Michael Vogt
University of Bonn

Oliver Linton
University of Cambridge

In this supplement, we provide the technical details omitted in the paper. In Section S.1, we prove Proposition 2.1 which concerns identification of the functions m_i. Sections S.2 and S.3 contain some auxiliary results needed for the proof of Theorem 6.1. In Section S.2, we in particular derive a general uniform convergence result which is applied to the kernel smoothers $\widehat{m}_{i,h}$ in Section S.3. The final Section S.4 contains the proof of Theorem 6.1. Throughout the supplement, we use the following notation: The symbol C denotes a universal real constant which may take a different value on each occurrence. In addition, the symbols $C_0, C_1, \ldots$ are used to denote specific real constants that are defined in the course of the supplement. Unless stated differently, the constants $C, C_0, C_1, \ldots$ depend neither on the dimensions n and T, nor on the indices $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, T\}$, nor on the location-bandwidth points $(x, h) \in \mathcal{G}_T$. To emphasize that the constants $C, C_0, C_1, \ldots$ do not depend on any of these parameters, we refer to them as absolute constants in many places.

S.1 Proof of Proposition 2.1

Let $\overline{Y}_i$, $\overline{Y}_t^{(i)}$ and $\overline{\overline{Y}}^{(i)}$ be the sample averages introduced in (3.1), that is,

$$\overline{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}, \quad \overline{Y}_t^{(i)} = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n Y_{jt} \quad \text{and} \quad \overline{\overline{Y}}^{(i)} = \frac{1}{(n-1)T} \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{t=1}^T Y_{jt}.$$

Define $\overline{\varepsilon}_i$, $\overline{\varepsilon}_t^{(i)}$ and $\overline{\overline{\varepsilon}}^{(i)}$ analogously and set $\overline{m}_i = T^{-1} \sum_{t=1}^T m_i(X_{it})$, $\overline{m}_t^{(i)} = (n-1)^{-1} \sum_{j=1, j\neq i}^n m_j(X_{jt})$ and $\overline{\overline{m}}^{(i)} = (\{n-1\}T)^{-1} \sum_{j=1, j \neq i}^n \sum_{t=1}^T m_j(X_{jt})$. Straightforward calculations yield that

$$\begin{aligned} Y_{it} - \overline{Y}_i - \overline{Y}_t^{(i)} + \overline{\overline{Y}}^{(i)} &= m_i(X_{it}) - \overline{m}_i - \overline{m}_t^{(i)} + \overline{\overline{m}}^{(i)} \\ &\quad + \varepsilon_{it} - \overline{\varepsilon}_i - \overline{\varepsilon}_t^{(i)} + \overline{\overline{\varepsilon}}^{(i)}. \end{aligned} \tag{S.1}$$

Hence, by adding/subtracting the sample averages $\overline{Y}_i$, $\overline{Y}_t^{(i)}$ and $\overline{\overline{Y}}^{(i)}$ from Y_{it}, we can eliminate the fixed effects α_i and γ_t from the model equation (2.5). We now consider the transformed model equation (S.1) for arbitrary but fixed indices i and t and examine the following two cases separately: (a) $n = n(T) \to \infty$ as $T \to \infty$, and (b) $n = n(T)$ remains bounded as $T \to \infty$.

(a) Under the normalization constraint (2.6) and the assumptions of Proposition 2.1, it holds that for any fixed i and t, $\overline{\varepsilon}_i = O_p(T^{-1/2})$ and $\overline{m}_i = O_p(T^{-1/2})$, $\overline{\varepsilon}_t^{(i)} = O_p(n^{-1/2})$ and $\overline{m}_t^{(i)} = O_p(n^{-1/2})$ as well as $\overline{\overline{\varepsilon}}^{(i)} = O_p(\{nT\}^{-1/2})$ and $\overline{\overline{m}}^{(i)} = O_p(\{nT\}^{-1/2})$. Using these facts in equation (S.1) for a fixed pair of indices i and t, we obtain that

$$Y_{it}^{\infty} = m_i(X_{it}) + \varepsilon_{it} \quad \text{a.s.}, \tag{S.2}$$

where Y_{it}^{∞} denotes the limit of $\widehat{Y}_{it}^{*} = Y_{it} - \overline{Y}_i - \overline{Y}_t^{(i)} + \overline{\overline{Y}}^{(i)}$ in probability, that is, $\widehat{Y}_{it}^{*} \xrightarrow{P} Y_{it}^{\infty}$. From (S.2), it follows that $\mathbb{E}[Y_{it}^{\infty}|X_{it}] = m_i(X_{it})$ almost surely, which identifies m_i.

(b) Now suppose that $n = n(T)$ remains bounded as $T \to \infty$. Let us assume for simplicity that $n = n(T)$ is non-decreasing in T, implying that n is a fixed number for sufficiently large T. (Without this assumption, we would have to consider a subsequence of time series lengths T_k for $k = 1, 2, \ldots$ such that $n(T_k)$ is non-decreasing.) Similar to the previous case, we have that $\overline{\varepsilon}_i = O_p(T^{-1/2})$ and $\overline{m}_i = O_p(T^{-1/2})$ as well as $\overline{\overline{\varepsilon}}^{(i)} = O_p(T^{-1/2})$ and $\overline{\overline{m}}^{(i)} = O_p(T^{-1/2})$. Using these facts in equation (S.1), we arrive at

$$Y_{it}^{\infty} = m_i(X_{it}) + \varepsilon_{it} - \big\{\overline{m}_t^{(i)} + \overline{\varepsilon}_t^{(i)}\big\} \quad \text{a.s.}, \tag{S.3}$$

where Y_{it}^{∞} is defined as before and, slightly abusing notation, we let $\overline{\varepsilon}_t^{(i)} = (N-1)^{-1}\sum_{j=1,j\neq i}^{N} \varepsilon_{jt}$ and $\overline{m}_t^{(i)} = (N-1)^{-1}\sum_{j=1,j\neq i}^{N} m_j(X_{jt})$ with $N = \lim_{T\to\infty} n(T)$. Since $\mathbb{E}[\overline{\varepsilon}_t^{(i)}|X_{it}] = \mathbb{E}[\overline{\varepsilon}_t^{(i)}] = 0$ and $\mathbb{E}[\overline{m}_t^{(i)}|X_{it}] = \mathbb{E}[\overline{m}_t^{(i)}] = 0$ under the normalization constraint (2.6) and the assumptions of Proposition 2.1, we get that $\mathbb{E}[Y_{it}^{\infty}|X_{it}] = m_i(X_{it})$ almost surely, which once again identifies m_i.

S.2 A general result on uniform convergence

In this and the subsequent section, we derive some uniform convergence results needed for the proof of Theorem 6.1. The multiscale statistics $\widehat{d}_{ij}$ are composed of kernel estimators whose building blocks are kernel averages of the form

$$\Phi_i(x,h) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)\Big(\frac{X_{it}-x}{h}\Big)^{\ell} Z_{it,T}, \tag{S.4}$$

where ℓ is a fixed natural number and X_{it} are the regressor variables from model (2.1). Moreover, $Z_{it,T}$ are general real-valued random variables that may depend on the sample size parameter T. For each i, the variables $(Z_{it,T}, X_{it})$ form a triangular array $\mathcal{A}_i = \{\mathcal{A}_{i,T}\}_{T=1}^{\infty}$, where $\mathcal{A}_{i,T} = \{(Z_{it,T}, X_{it}) : 1 \le t \le T\}$. We make the following assumptions on the random variables $(Z_{it,T}, X_{it})$:

(P1) For each i and T, the collection of random variables $\mathcal{A}_{i,T}$ is strongly mixing. The mixing coefficients $\alpha_{i,T}(\ell)$ of $\mathcal{A}_{i,T}$ are such that $\alpha_{i,T}(\ell) \le n\,\alpha(\ell)$ for all i, T and ℓ, where the coefficients $\alpha(\ell)$ decay exponentially fast to zero as $\ell \to \infty$.

(P2) There exist a real number $\theta > 2$ and a natural number ℓ^* such that for any $\ell \in \mathbb{Z}$ with $|\ell| \ge \ell^*$ and some absolute constant $C < \infty$,

$$\max_{1\le t\le T} \max_{1\le i\le n} \sup_{x\in[0,1]} \mathbb{E}\big[|Z_{it,T}|^\theta \big| X_{it} = x\big] \le C < \infty$$
$$\max_{1\le t\le T} \max_{1\le i\le n} \sup_{x,x'\in[0,1]} \mathbb{E}\big[|Z_{it,T} Z_{it+\ell,T}| \big| X_{it} = x, X_{it+\ell} = x'\big] \le C < \infty.$$

The following lemma characterizes the convergence behaviour of the kernel average $\Phi_i(x,h)$ uniformly over i, x and h.

Proposition S.1. *Let (P1) and (P2) be satisfied. Moreover, assume that (C2) and (C7)–(C9) are fulfilled. Then it holds that*

$$\mathbb{P}\Big(\max_{1\le i\le n} \max_{(x,h)\in\mathcal{G}_T} \sqrt{Th}\big|\Phi_i(x,h) - \mathbb{E}\Phi_i(x,h)\big| > C_0\sqrt{\gamma_{n,T}}\Big) = o(1),$$

where $\gamma_{n,T} = \log n + \log T$ and C_0 is a sufficiently large absolute constant.

Proof of Proposition S.1. To prove the proposition, we modify standard arguments to derive uniform convergence rates for kernel estimators, which can be found e.g. in Masry (1996), Bosq (1998) or Hansen (2008). These arguments were originally designed to derive the convergence rates of kernel averages such as $\Phi_i(x,h) - \mathbb{E}\Phi_i(x,h)$ uniformly over x but pointwise in h and i. In contrast to this, we aim to derive the convergence rate of $\Phi_i(x,h) - \mathbb{E}\Phi_i(x,h)$ uniformly over x, h and i. Related results can be found e.g. in Einmahl and Mason (2005) and Vogt and Linton (2017) (see in particular Lemma S.1 therein).

We now turn to the proof of the proposition. For simplicity of notation, we let $\ell = 0$ in (S.4), the arguments being completely analogous for $\ell \ne 0$. To start with, we define

$$Z^{\le}_{it,T} = Z_{it,T}\, 1\big(|Z_{it,T}| \le (nT)^{\frac{1}{\theta-\delta}}\big)$$
$$Z^{>}_{it,T} = Z_{it,T}\, 1\big(|Z_{it,T}| > (nT)^{\frac{1}{\theta-\delta}}\big),$$

where $\delta > 0$ is an absolute constant that can be chosen as small as desired. Moreover, we write

$$\sqrt{Th}\big\{\Phi_i(x,h) - \mathbb{E}\Phi_i(x,h)\big\} = \sum_{t=1}^T \mathcal{Z}^{\le}_{it,T}(x,h) + \sum_{t=1}^T \mathcal{Z}^{>}_{it,T}(x,h),$$

where

$$\mathcal{Z}_{it,T}^{\leq}(x,h) = \frac{1}{\sqrt{Th}}\Big\{K\Big(\frac{X_{it}-x}{h}\Big)Z_{it,T}^{\leq} - \mathbb{E}\Big[K\Big(\frac{X_{it}-x}{h}\Big)Z_{it,T}^{\leq}\Big]\Big\}$$
$$\mathcal{Z}_{it,T}^{>}(x,h) = \frac{1}{\sqrt{Th}}\Big\{K\Big(\frac{X_{it}-x}{h}\Big)Z_{it,T}^{>} - \mathbb{E}\Big[K\Big(\frac{X_{it}-x}{h}\Big)Z_{it,T}^{>}\Big]\Big\}.$$

With this notation at hand, we get that

$$\mathbb{P}\Big(\max_{1\leq i\leq n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\big|\Phi_i(x,h)-\mathbb{E}\Phi_i(x,h)\big| > C_0\sqrt{\gamma_{n,T}}\Big) \leq P^{\leq} + P^{>},$$

where

$$P^{\leq} = \mathbb{P}\Big(\max_{1\leq i\leq n}\max_{(x,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T}\mathcal{Z}_{it,T}^{\leq}(x,h)\Big| > \frac{C_0}{2}\sqrt{\gamma_{n,T}}\Big)$$
$$P^{>} = \mathbb{P}\Big(\max_{1\leq i\leq n}\max_{(x,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T}\mathcal{Z}_{it,T}^{>}(x,h)\Big| > \frac{C_0}{2}\sqrt{\gamma_{n,T}}\Big).$$

In what follows, we show that $P^{\leq} = o(1)$ and $P^{>} = o(1)$, which implies the statement of Proposition S.1.

We first have a closer look at $P^{>}$. It holds that

$$P^{>} \leq \sum_{i=1}^{n}\mathbb{P}\Big(\max_{(x,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T}\mathcal{Z}_{it,T}^{>}(x,h)\Big| > \frac{C_0}{2}\sqrt{\gamma_{n,T}}\Big) \leq P_1^{>} + P_2^{>},$$

where

$$P_1^{>} = \sum_{i=1}^{n}\mathbb{P}\Big(\max_{(x,h)\in\mathcal{G}_T}\Big|\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\Big(\frac{X_{it}-x}{h}\Big)Z_{it,T}^{>}\Big| > \frac{C_0}{4}\sqrt{\gamma_{n,T}}\Big)$$
$$P_2^{>} = \sum_{i=1}^{n}\mathbb{P}\Big(\max_{(x,h)\in\mathcal{G}_T}\Big|\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\mathbb{E}\Big[K\Big(\frac{X_{it}-x}{h}\Big)Z_{it,T}^{>}\Big]\Big| > \frac{C_0}{4}\sqrt{\gamma_{n,T}}\Big).$$

With the help of (P2), we obtain that

$$\begin{aligned}
P_1^{>} &\leq \sum_{i=1}^{n}\mathbb{P}\Big(|Z_{it,T}| > (nT)^{\frac{1}{\theta-\delta}} \text{ for some } 1\leq t\leq T\Big)\\
&\leq \sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{P}\Big(|Z_{it,T}| > (nT)^{\frac{1}{\theta-\delta}}\Big)\\
&\leq C(nT)\big/(nT)^{\frac{\theta}{\theta-\delta}}\\
&= o(1).
\end{aligned}$$

Once again exploiting (P2), we can further infer that

$$\begin{aligned}\Big|\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\mathbb{E}\Big[K\Big(\frac{X_{it}-x}{h}\Big)Z_{it,T}^{>}\Big]\Big| &\le \frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\mathbb{E}\Big[K\Big(\frac{X_{it}-x}{h}\Big)\frac{|Z_{it,T}|^{\theta}}{(nT)^{\frac{\theta-1}{\theta-\delta}}}\Big] \\ &\le C\sqrt{Th}/(nT)^{\frac{\theta-1}{\theta-\delta}} \\ &= o\big(\sqrt{\gamma_{n,T}}\big),\end{aligned}$$

which immediately implies that $P_2^{>} = 0$ for sufficiently large T. Putting everything together, we arrive at the result that $P^{>} = o(1)$.

We now turn to the analysis of $P^{\le}$. In what follows, we show that

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\mathbb{P}\Big(\Big|\sum_{t=1}^{T}\mathcal{Z}_{it,T}^{\le}(x,h)\Big| > \frac{C_0}{2}\sqrt{\gamma_{n,T}}\Big) \le CT^{-r}, \tag{S.5}$$

where the constant $r > 0$ can be chosen as large as desired. From (S.5), it immediately follows that $P^{\le} = o(1)$, since

$$P^{\le} \le \sum_{i=1}^{n}\sum_{(x,h)\in\mathcal{G}_T}\mathbb{P}\Big(\Big|\sum_{t=1}^{T}\mathcal{Z}_{it,T}^{\le}(x,h)\Big| > \frac{C_0}{2}\sqrt{\gamma_{n,T}}\Big).$$

To complete the proof of Proposition S.1, it thus remains to verify (S.5). To do so, we split the term $\sum_{t=1}^{T}\mathcal{Z}_{it,T}^{\le}(x,h)$ into blocks as follows:

$$\sum_{t=1}^{T}\mathcal{Z}_{it,T}^{\le}(x,h) = \sum_{s=1}^{\lceil N_T\rceil}B_{2s-1} + \sum_{s=1}^{\lfloor N_T\rfloor}B_{2s}$$

with

$$B_s = B_{is}(x,h) = \sum_{t=(s-1)L_T+1}^{\min\{sL_T,T\}}\mathcal{Z}_{it,T}^{\le}(x,h),$$

where $L_T = L_{T,h} = \sqrt{Th/\gamma_{n,T}}\,(nT)^{-1/(\theta-\delta)}$ is the block length and $2N_T$ with $N_T = \lceil T/L_T\rceil/2$ is the number of blocks. Note that under condition (6.1), it holds that $cT^{\xi} \le L_{T,h} \le CT^{1-\xi}$ for any h with $h_{\min} \le h \le h_{\max}$ and some sufficiently small $\xi > 0$, where c, C and ξ are absolute constants that in particular do not depend on h. With this notation at hand, we obtain that

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{t=1}^{T}\mathcal{Z}_{it,T}^{\le}(x,h)\Big| > \frac{C_0}{2}\sqrt{\gamma_{n,T}}\Big) &\le \mathbb{P}\Big(\Big|\sum_{s=1}^{\lceil N_T\rceil}B_{2s-1}\Big| > \frac{C_0}{4}\sqrt{\gamma_{n,T}}\Big) \\ &\quad + \mathbb{P}\Big(\Big|\sum_{s=1}^{\lfloor N_T\rfloor}B_{2s}\Big| > \frac{C_0}{4}\sqrt{\gamma_{n,T}}\Big).\end{aligned} \tag{S.6}$$

As the two terms on the right-hand side of (S.6) can be treated analogously, we focus attention to the first one. By Bradley's strong approximation theorem (see Theorem 3 in Bradley (1983)), we can construct a sequence of random variables $B_1^*, B_3^*, \dots$ such that (i) $B_1^*, B_3^*, \dots$ are independent, (ii) B_{2s-1} and B_{2s-1}^* have the same distribution for each s, and (iii) for $0 < \mu \le \|B_{2s-1}\|_\infty$, $\mathbb{P}(|B_{2s-1}^* - B_{2s-1}| > \mu) \le 18(\|B_{2s-1}\|_\infty/\mu)^{1/2} n\, \alpha(L_T)$. With the variables B_{2s-1}^*, we can construct the bound

$$\mathbb{P}\Big(\Big| \sum_{s=1}^{\lceil N_T \rceil} B_{2s-1} \Big| > \frac{C_0}{4} \sqrt{\gamma_{n,T}} \Big) \le P_1^* + P_2^*, \tag{S.7}$$

where

$$\begin{aligned} P_1^* &= \mathbb{P}\Big(\Big| \sum_{s=1}^{\lceil N_T \rceil} B_{2s-1}^* \Big| > \frac{C_0}{8} \sqrt{\gamma_{n,T}} \Big) \\ P_2^* &= \mathbb{P}\Big(\Big| \sum_{s=1}^{\lceil N_T \rceil} \big(B_{2s-1} - B_{2s-1}^*\big) \Big| > \frac{C_0}{8} \sqrt{\gamma_{n,T}} \Big). \end{aligned}$$

Using (iii) together with the fact that the mixing coefficients $\alpha(\cdot)$ decay to zero exponentially fast, it is not difficult to see that $P_2^* \le CT^{-r}$, where the constant $r > 0$ can be picked as large as desired. To deal with P_1^*, we make use of the following three facts:

(a) For a real-valued random variable B and $\lambda > 0$, Markov's inequality yields that $\mathbb{P}(\pm B > \delta) \le \mathbb{E} \exp(\pm \lambda B) / \exp(\lambda \delta)$.

(b) Since $|B_{2s-1}| \le \{C L_T (nT)^{1/(\theta-\delta)}\}/\sqrt{Th}$, it holds that $\lambda_{n,T} |B_{2s-1}| \le 1/2$, where we set $\lambda_{n,T} = \sqrt{Th}/\{2 C L_T (nT)^{1/(\theta-\delta)}\}$. As $\exp(x) \le 1 + x + x^2$ for $|x| \le 1/2$, we get that

$$\mathbb{E}\Big[\exp\big(\pm \lambda_{n,T} B_{2s-1} \big) \Big] \le 1 + \lambda_{n,T}^2 \mathbb{E}\big[(B_{2s-1})^2 \big] \le \exp\big(\lambda_{n,T}^2 \mathbb{E}\big[(B_{2s-1})^2 \big] \big)$$

along with

$$\mathbb{E}\Big[\exp\big(\pm \lambda_{n,T} B_{2s-1}^* \big) \Big] \le \exp\big(\lambda_{n,T}^2 \mathbb{E}\big[(B_{2s-1}^*)^2 \big] \big).$$

(c) Standard calculations for kernel estimators yield that $\sum_{s=1}^{\lceil N_T \rceil} \mathbb{E}\big[(B_{2s-1}^*)^2 \big] \le C_2$.

Using (a)–(c), we obtain that

$$P_1^* \le \mathbb{P}\Big(\sum_{s=1}^{\lceil N_T \rceil} B_{2s-1}^* > \frac{C_0}{8} \sqrt{\gamma_{n,T}} \Big) + \mathbb{P}\Big(- \sum_{s=1}^{\lceil N_T \rceil} B_{2s-1}^* > \frac{C_0}{8} \sqrt{\gamma_{n,T}} \Big),$$

where

$$
\begin{aligned}
&\mathbb{P}\Big(\pm \sum_{s=1}^{\lceil N_T \rceil} B^*_{2s-1} > \frac{C_0}{8}\sqrt{\gamma_{n,T}}\Big) \\
&\le \exp\Big(-\frac{C_0}{8}\lambda_{n,T}\sqrt{\gamma_{n,T}}\Big)\, \mathbb{E}\Big[\exp\Big(\pm \lambda_{n,T} \sum_{s=1}^{\lceil N_T \rceil} B^*_{2s-1}\Big)\Big] \\
&\le \exp\Big(-\frac{C_0}{8}\lambda_{n,T}\sqrt{\gamma_{n,T}}\Big) \prod_{s=1}^{\lceil N_T \rceil} \mathbb{E}\Big[\exp\big(\pm \lambda_{n,T} B^*_{2s-1}\big)\Big] \\
&\le \exp\Big(-\frac{C_0}{8}\lambda_{n,T}\sqrt{\gamma_{n,T}}\Big) \prod_{s=1}^{\lceil N_T \rceil} \exp\Big(\lambda^2_{n,T}\mathbb{E}\big[(B^*_{2s-1})^2\big]\Big) \\
&= \exp\Big(-\frac{C_0}{8}\lambda_{n,T}\sqrt{\gamma_{n,T}}\Big) \exp\Big(\lambda^2_{n,T} \sum_{s=1}^{\lceil N_T \rceil} \mathbb{E}\big[(B^*_{2s-1})^2\big]\Big) \\
&\le \exp\Big(-\frac{C_0}{8}\lambda_{n,T}\sqrt{\gamma_{n,T}} + C_2\lambda^2_{n,T}\Big).
\end{aligned}
$$

From the definition of $\lambda_{n,T}$, it follows that $\lambda_{n,T} = C_3\sqrt{\gamma_{n,T}}$ with some absolute constant $C_3 > 0$. Hence,

$$
\begin{aligned}
P^*_1 &\le 2\exp\Big(-\frac{C_0}{8}\lambda_{n,T}\sqrt{\gamma_{n,T}} + C_2\lambda^2_{n,T}\Big) \\
&= 2\exp\Big(-\frac{C_0C_3}{8}\{\log n + \log T\} + C_2C_3^2\{\log n + \log T\}\Big) \le CT^{-r},
\end{aligned}
$$

where the constant $r > 0$ can be made arbitrarily large by picking C_0 large enough. To summarize, we have shown that $P^*_1 \le CT^{-r}$ and $P^*_2 \le CT^{-r}$ with some arbitrarily large $r > 0$. This together with the bounds from (S.7) and (S.6) yields (S.5), which in turn completes the proof. □

S.3 Auxiliary results on uniform convergence

We now use Proposition S.1 from the previous section to derive the uniform convergence rates of some kernel estimators of interest. To start with, we consider the kernel averages

$$
S_{i,\ell}(x,h) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)\Big(\frac{X_{it}-x}{h}\Big)^{\ell} \tag{S.8}
$$

$$
S^+_{i,\ell}(x,h) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)\Big|\frac{X_{it}-x}{h}\Big|^{\ell} \tag{S.9}
$$

$$
S^{\varepsilon}_{i,\ell}(x,h) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)\Big(\frac{X_{it}-x}{h}\Big)^{\ell}\varepsilon_{it} \tag{S.10}
$$

$$S^m_{i,\ell}(x,h) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)\Big(\frac{X_{it}-x}{h}\Big)^{\ell}\{m_i(X_{it})-m_i(x)\} \tag{S.11}$$

for $0 \le \ell \le 3$.

Lemma S.2. *Under (C1), (C2) and (C5)–(C9), it holds that*

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T} \sqrt{Th}\,\big|S_{i,\ell}(x,h) - \mathbb{E}[S_{i,\ell}(x,h)]\big| = O_p\big(\sqrt{\gamma_{n,T}}\big) \tag{S.12}$$

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T} \sqrt{Th}\,\big|S^+_{i,\ell}(x,h) - \mathbb{E}[S^+_{i,\ell}(x,h)]\big| = O_p\big(\sqrt{\gamma_{n,T}}\big) \tag{S.13}$$

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T} \sqrt{Th}\,\big|S^\varepsilon_{i,\ell}(x,h)\big| = O_p\big(\sqrt{\gamma_{n,T}}\big) \tag{S.14}$$

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T} \sqrt{Th}\,\big|S^m_{i,\ell}(x,h) - \mathbb{E}[S^m_{i,\ell}(x,h)]\big| = O_p\big(\sqrt{\gamma_{n,T}}\big) \tag{S.15}$$

with $\gamma_{n,T} = \log n + \log T$.

Proof of Lemma S.2. The terms $S_{i,\ell}(x,h)$ and $S^\varepsilon_{i,\ell}(x,h)$ can be written in the form $T^{-1}\sum_{t=1}^T K_h(X_{it}-x)\{(X_{it}-x)/h\}^\ell Z_{it,T}$ with $Z_{it,T}=1$ and $Z_{it,T}=\varepsilon_{it}$, respectively. In addition, $S^m_{i,\ell}(x,h)$ can be expressed as $S^m_{i,\ell}(x,h) = T^{-1}\sum_{t=1}^T K_h(X_{it}-x)\{(X_{it}-x)/h\}^\ell Z^A_{it,T} - m_i(x)\,T^{-1}\sum_{t=1}^T K_h(X_{it}-x)\{(X_{it}-x)/h\}^\ell Z^B_{it,T}$ with $Z^A_{it,T}=m_i(X_{it})$ and $Z^B_{it,T}=1$. Hence, the statements (S.12), (S.14) and (S.15) are simple consequences of Proposition S.1. Moreover, it is trivial to modify the proof of Proposition S.1 to apply to the expression $S^+_{i,\ell}(x,h)$ and thus to derive statement (S.13). $\square$

The terms $S_{i,\ell}(x,h)$, $S^\varepsilon_{i,\ell}(x,h)$ and $S^m_{i,\ell}(x,h)$ are the building blocks of the local linear kernel averages

$$Q_i(x,h) = \frac{1}{T}\sum_{t=1}^T W_{it}(x,h) \tag{S.16}$$

$$Q^\varepsilon_i(x,h) = \frac{1}{T}\sum_{t=1}^T W_{it}(x,h)\varepsilon_{it} \tag{S.17}$$

$$Q^m_i(x,h) = \frac{1}{T}\sum_{t=1}^T W_{it}(x,h)\{m_i(X_{it})-m_i(x)\}. \tag{S.18}$$

In particular, it holds that

$$\begin{aligned}
Q_i(x,h) &= S_{i,2}(x,h)S_{i,0}(x,h) - S^2_{i,1}(x,h)\\
Q^\varepsilon_i(x,h) &= S_{i,2}(x,h)S^\varepsilon_{i,0}(x,h) - S_{i,1}(x,h)S^\varepsilon_{i,1}(x,h)\\
Q^m_i(x,h) &= S_{i,2}(x,h)S^m_{i,0}(x,h) - S_{i,1}(x,h)S^m_{i,1}(x,h).
\end{aligned}$$

The uniform convergence rates of $Q_i(x,h)$, $Q^\varepsilon_i(x,h)$ and $Q^m_i(x,h)$ can be easily derived with the help of Lemma S.2 and some additional straightforward arguments. Defining

$$
\begin{aligned}
Q_i^*(x,h) &= \mathbb{E}[S_{i,2}(x,h)]\mathbb{E}[S_{i,0}(x,h)] - \mathbb{E}[S_{i,1}(x,h)]^2 \\
Q_i^{m,*}(x,h) &= \mathbb{E}[S_{i,2}(x,h)]\mathbb{E}[S_{i,0}^m(x,h)] - \mathbb{E}[S_{i,1}(x,h)]\mathbb{E}[S_{i,1}^m(x,h)],
\end{aligned}
$$

we in particular obtain the following result.

Lemma S.3. *Under (C1), (C2) and (C5)–(C9), it holds that*

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\,\big|Q_i(x,h)-Q_i^*(x,h)\big| = O_p\big(\sqrt{\gamma_{n,T}}\big) \tag{S.19}$$

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\,\big|Q_i^{\varepsilon}(x,h)\big| = O_p\big(\sqrt{\gamma_{n,T}}\big) \tag{S.20}$$

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\,\big|Q_i^m(x,h)-Q_i^{m,*}(x,h)\big| = O_p\big(\sqrt{\gamma_{n,T}}\big) \tag{S.21}$$

with $\gamma_{n,T} = \log n + \log T$.

In addition to $Q_i(x,h)$, $Q_i^{\varepsilon}(x,h)$ and $Q_i^m(x,h)$, we consider the kernel average

$$Q_i^{\mathrm{fe}}(x,h) = \frac{1}{T}\sum_{t=1}^T W_{it}(x,h)\big\{\overline{\varepsilon}_t^{(i)} + \overline{m}_t^{(i)}\big\},$$

whose uniform convergence rate is specified by the following lemma.

Lemma S.4. *Under (C1), (C2) and (C5)–(C9), it holds that*

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\,\big|Q_i^{\mathrm{fe}}(x,h)\big| = O_p\big(\sqrt{\log n+\log T}\big).$$

Proof of Lemma S.4. Defining

$$S_{i,\ell}^{\mathrm{fe}}(x,h) = \frac{1}{T}\sum_{t=1}^T K_h(X_{it}-x)\Big(\frac{X_{it}-x}{h}\Big)^{\ell} Z_{it,T}$$

with $Z_{it,T} = \overline{\varepsilon}_t^{(i)} + \overline{m}_t^{(i)}$, we can write $Q_i^{\mathrm{fe}}(x,h) = S_{i,2}(x,h)S_{i,0}^{\mathrm{fe}}(x,h) - S_{i,1}(x,h)S_{i,1}^{\mathrm{fe}}(x,h)$. From (C1) and Theorem 5.1(a) in Bradley (2005), it follows that the collection of random variables $\mathcal{A}_{i,T} = \{(X_{it}, Z_{it,T}) : 1 \le t \le T\}$ is strongly mixing for any i and T. In particular, the mixing coefficients $\alpha_{i,T}(\ell)$ of $\mathcal{A}_{i,T}$ are such that $\alpha_{i,T}(\ell) \le n\,\alpha(\ell)$, where the coefficients $\alpha(\ell)$ are defined in (C1) and decay exponentially fast to zero. According to this, the variables $(Z_{it,T}, X_{it})$ satisfy condition (P1). Since the collection of random variables $\{Z_{it,T} : 1 \le t \le T\}$ is independent from $\{X_{it} : 1 \le t \le T\}$ for any i under (C1), it is straightforward to verify that the variables $(Z_{it,T}, X_{it})$ fulfill condition (P2) as well. Hence, we can apply Proposition S.1 to get that

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\,\big|S_{i,\ell}^{\mathrm{fe}}(x,h)\big| = O_p\big(\sqrt{\log n+\log T}\big).$$

With this and Lemma S.2, it is straightforward to complete the proof. $\square$

With the help of the kernel averages defined and analyzed above, the local linear kernel smoothers $\widehat{m}_{i,h}$ can be expressed as

$$\widehat{m}_{i,h}(x) - m_i(x) = \frac{Q_i^{\varepsilon}(x,h) + Q_i^{m}(x,h) - Q_i^{\mathrm{fe}}(x,h)}{Q_i(x,h)} - \big\{\overline{m}_i + \overline{\varepsilon}_i\big\} + \big\{\overline{\overline{m}}^{(i)} + \overline{\overline{\varepsilon}}^{(i)}\big\}.$$

We now use this formulation to derive two different uniform expansions of the term $\sqrt{Th}\{\widehat{m}_{i,h}(x) - m_i(x)\}$, which are required to prove different parts of Theorem 6.1.

Proposition S.5. *Let the conditions of Theorem 6.1 be satisfied. Then it holds that*

$$\sqrt{Th}\big\{\widehat{m}_{i,h}(x) - m_i(x)\big\} = \sqrt{Th}\frac{Q_i^{m,*}(x,h)}{Q_i^*(x,h)} + R_i^{(a)}(x,h),$$

where the remainder $R_i^{(a)}(x,h)$ has the property that

$$\max_{1 \le i \le n} \max_{(x,h) \in \mathcal{G}_T} \big|R_i^{(a)}(x,h)\big| = O_p\big(\sqrt{\log n + \log T}\big).$$

Proposition S.6. *Under the conditions of Theorem 6.1, it holds that*

$$\sqrt{Th}\big\{\widehat{m}_{i,h}(x) - m_i(x)\big\} = \sqrt{Th^5}\frac{\kappa(x,h)m_i''(x)}{2} + R_i^{(b)}(x,h),$$

where we use the shorthand $\kappa(x,h) = \{\kappa_2(x,h)^2 - \kappa_1(x,h)\kappa_3(x,h)\}/\{\kappa_2(x,h)\kappa_0(x,h) - \kappa_1(x,h)^2\}$ with $\kappa_\ell(x,h) = \int_{-x/h}^{(1-x)/h} u^\ell K(u)du$ and the remainder $R_i^{(b)}(x,h)$ is such that

$$\max_{1 \le i \le n} \max_{(x,h) \in \mathcal{G}_T} \big|R_i^{(b)}(x,h)\big| = O_p\big(\sqrt{\log n + \log T} + \sqrt{Th_{\max}^7}\big).$$

Proof of Proposition S.5. Simple algebra yields that

$$\sqrt{Th}\big\{\widehat{m}_{i,h}(x) - m_i(x)\big\} = \sqrt{Th}\frac{Q_i^{m,*}(x,h)}{Q_i^*(x,h)} + R_i^{(a)}(x,h),$$

where $R_i^{(a)}(x,h) = R_{i,1}^{(a)}(x,h) + \ldots + R_{i,6}^{(a)}(x,h)$ with

$$\begin{aligned}
R_{i,1}^{(a)}(x,h) &= \sqrt{Th}Q_i^{m,*}(x,h)\Big\{\frac{1}{Q_i(x,h)} - \frac{1}{Q_i^*(x,h)}\Big\} \\
R_{i,2}^{(a)}(x,h) &= \sqrt{Th}\frac{Q_i^{m}(x,h) - Q_i^{m,*}(x,h)}{Q_i(x,h)} \\
R_{i,3}^{(a)}(x,h) &= \sqrt{Th}\frac{Q_i^{\varepsilon}(x,h)}{Q_i(x,h)} \\
R_{i,4}^{(a)}(x,h) &= -\sqrt{Th}\frac{Q_i^{\mathrm{fe}}(x,h)}{Q_i(x,h)}
\end{aligned}$$

as well as $R^{(a)}_{i,5}(x,h) = -\sqrt{Th}\{\overline{m}_i + \overline{\varepsilon}_i\}$ and $R^{(a)}_{i,6}(x,h) = \sqrt{Th}\{\overline{\overline{m}}^{(i)} + \overline{\overline{\varepsilon}}^{(i)}\}$. To complete the proof, we show that

$$\max_{1 \le i \le n} \max_{(x,h) \in \mathcal{G}_T} \big| R^{(a)}_{i,\ell}(x,h) \big| = O_p\big(\sqrt{\log n + \log T}\big) \tag{S.22}$$

for $1 \le \ell \le 6$: By standard bias calculations, we obtain that

$$\max_{1 \le i \le n} \max_{(x,h) \in \mathcal{G}_T} \big| Q^{m,*}_i(x,h) \big| = O(h_{\max}) \tag{S.23}$$

$$\max_{1 \le i \le n} \max_{(x,h) \in \mathcal{G}_T} \big| Q^{*}_i(x,h) - \big\{\kappa_2(x,h)\kappa_0(x,h) - \kappa_1(x,h)^2\big\} f_i^2(x) \big| = O(h_{\max}), \tag{S.24}$$

where under our assumptions, the term $Q^{**}_i(x,h) = \{\kappa_2(x,h)\kappa_0(x,h) - \kappa_1(x,h)^2\} f_i^2(x)$ is bounded away from zero and infinity uniformly over i and (x,h), that is, $0 < c \le Q^{**}_i(x,h) \le C < \infty$ with some constants c and C that are independent of i and (x,h). With the help of these observations and Lemmas S.3 and S.4, it is straightforward to derive (S.22) for $1 \le \ell \le 4$. Next, note that $\max_{1 \le i \le n} |\overline{\overline{m}}^{(i)}| \le \max_{1 \le i \le n} |\overline{m}_i|$ and $\max_{1 \le i \le n} |\overline{\overline{\varepsilon}}^{(i)}| \le \max_{1 \le i \le n} |\overline{\varepsilon}_i|$. Arguments similar to but simpler than those for Proposition S.1 yield that $\max_{1 \le i \le n} |\overline{m}_i| = O_p(\sqrt{\{\log n + \log T\}/T})$ and $\max_{1 \le i \le n} |\overline{\varepsilon}_i| = O_p(\sqrt{\{\log n + \log T\}/T})$. From this, (S.22) immediately follows for $\ell = 5$ and $\ell = 6$. $\square$

Proof of Proposition S.6. Straightforward calculations yield that

$$\sqrt{Th}\big\{\widehat{m}_{i,h}(x) - m_i(x)\big\} = \sqrt{Th^5}\frac{\kappa(x,h) m_i''(x)}{2} + R^{(b)}_i(x,h),$$

where $R^{(b)}_i(x,h) = R^{(b)}_{i,1}(x,h) + \ldots + R^{(b)}_{i,5}(x,h)$ with

$$R^{(b)}_{i,1}(x,h) = \sqrt{Th}\Big\{\frac{Q^m_i(x,h)}{Q_i(x,h)} - h^2 \frac{\kappa(x,h) m_i''(x)}{2}\Big\}$$

and $R^{(b)}_{i,\ell}(x,h) = R^{(a)}_{i,\ell+1}(x,h)$ for $2 \le \ell \le 5$. In order to prove Proposition S.6, it suffices to show that

$$\max_{1 \le i \le n} \max_{(x,h) \in \mathcal{G}_T} \big| R^{(b)}_{i,1}(x,h) \big| = O_p\big(\sqrt{Th^7_{\max}}\big) + o_p\big(\sqrt{\log n + \log T}\big) \tag{S.25}$$

$$\max_{1 \le i \le n} \max_{(x,h) \in \mathcal{G}_T} \big| R^{(b)}_{i,\ell}(x,h) \big| = O_p\big(\sqrt{\log n + \log T}\big) \tag{S.26}$$

for $2 \le \ell \le 5$. (S.26) has already been verified in the proof of Proposition S.5. To prove (S.25), we make use of the following two facts:

(a) From Lemma S.3 and (S.24), it follows that

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\big|Q_i(x,h)-Q_i^{**}(x,h)\big| = O_p\big(\sqrt{\log n+\log T}+\sqrt{Th_{\max}^3}\big) \quad (S.27)$$

with $Q_i^{**}(x,h)=\{\kappa_2(x,h)\kappa_0(x,h)-\kappa_1(x,h)^2\}f_i^2(x)$. As already noted in the proof of Proposition S.5, the term $Q_i^{**}(x,h)$ is bounded away from zero and infinity uniformly over i and (x,h).

(b) A second-order Taylor expansion of m_i yields that

$$\sqrt{Th}Q_i^m(x,h)=\sqrt{Th}Q_i^{m,**}(x,h)+R_i^m(x,h), \quad (S.28)$$

where

$$Q_i^{m,**}(x,h)=h^2\frac{m_i''(x)f_i^2(x)}{2}\big[\kappa_2(x,h)^2-\kappa_1(x,h)\kappa_3(x,h)\big].$$

The remainder term $R_i^m(x,h)$ has the form $R_i^m(x,h)=R_{i,1}^m(x,h)+R_{i,2}^m(x,h)$, where

$$\begin{aligned} R_{i,1}^m(x,h) &= \sqrt{Th^5}\frac{m_i''(x)}{2}\Big\{\big[S_{i,2}(x,h)^2-S_{i,1}(x,h)S_{i,3}(x,h)\big] \\ &\qquad -\big[\kappa_2(x,h)^2-\kappa_1(x,h)\kappa_3(x,h)\big]f_i^2(x)\Big\} \\ R_{i,2}^m(x,h) &= \frac{\sqrt{Th^5}}{2T}\sum_{t=1}^T K_h(X_{it}-x)\Big[S_{i,2}(x,h)-\Big(\frac{X_{it}-x}{h}\Big)S_{i,1}(x,h)\Big] \\ &\qquad \times\big\{m_i''(\xi_{it})-m_i''(x)\big\}\Big(\frac{X_{it}-x}{h}\Big)^2 \end{aligned}$$

with ξ_{it} denoting an intermediate point between X_{it} and x. By Lemma S.2 and standard bias calculations, we obtain that

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\big|R_{i,1}^m(x,h)\big| = O_p\big(h_{\max}^2\sqrt{\log n+\log T}+\sqrt{Th_{\max}^7}\big). \quad (S.29)$$

As m_i'' is Lipschitz continuous by (C6), we further get that $|R_{i,2}^m(x,h)|\le C\sqrt{Th^7}$ $\{S_{i,2}(x,h)^2+S_{i,1}^+(x,h)S_{i,3}^+(x,h)\}$. Applying Lemma S.2 together with standard bias calculations to this upper bound, we can infer that

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\big|R_{i,2}^m(x,h)\big| = O_p\big(h_{\max}^3\sqrt{\log n+\log T}+\sqrt{Th_{\max}^7}\big). \quad (S.30)$$

Finally, by combining (S.29) and (S.30), the remainder term $R_i^m(x,h)$ is seen to have the property that

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\big|R_i^m(x,h)\big| = O_p\big(h_{\max}^2\sqrt{\log n+\log T}+\sqrt{Th_{\max}^7}\big). \quad (S.31)$$

We now proceed as follows: Simple algebra yields that

$$\begin{aligned}
\sqrt{Th}&\Big(\frac{Q_i^m(x,h)}{Q_i(x,h)} - \frac{Q_i^{m,**}(x,h)}{Q_i^{**}(x,h)}\Big) \\
&= \frac{R_i^m(x,h)}{Q_i(x,h)} + \sqrt{Th}Q_i^{m,**}(x,h)\Big\{\frac{1}{Q_i(x,h)} - \frac{1}{Q_i^{**}(x,h)}\Big\}.
\end{aligned}$$

Since $Q_i^{m,**}(x,h)/Q_i^{**}(x,h) = h^2\kappa(x,h)m_i''(x)/2$, this implies that

$$R_{i,1}^{(b)}(x,h) = \frac{R_i^m(x,h)}{Q_i(x,h)} + \sqrt{Th}Q_i^{m,**}(x,h)\Big\{\frac{1}{Q_i(x,h)} - \frac{1}{Q_i^{**}(x,h)}\Big\}.$$

Using this representation of $R_{i,1}^{(b)}(x,h)$ together with (S.27), (S.31) and the fact that $Q_i^{**}(x,h)$ is bounded away from zero and infinity uniformly over i and (x,h), it is straightforward to verify (S.25). $\square$

The final result of this section is concerned with the normalization term

$$\widehat{\nu}_{ij}(x,h) = \left\{\frac{\widehat{\sigma}_{i,h}^2}{\widehat{f}_{i,h}(x)} + \frac{\widehat{\sigma}_{j,h}^2}{\widehat{f}_{j,h}(x)}\right\} s(x,h), \qquad \text{(S.32)}$$

where $s(x,h) = \{\int_{-x/h}^{(1-x)/h} K^2(u)[\kappa_2(x,h) - \kappa_1(x,h)u]^2 du\}/\{\kappa_0(x,h)\kappa_2(x,h) - \kappa_1(x,h)^2\}^2$ with $\kappa_\ell(x,h) = \int_{-x/h}^{(1-x)/h} u^\ell K(u)du$ for $0 \le \ell \le 2$, $\widehat{f}_{i,h}(x) = \{\kappa_0(x,h)T\}^{-1}\sum_{t=1}^T K_h(X_{it} - x)$ and $\widehat{\sigma}_{i,h}^2 = T^{-1}\sum_{t=1}^T\{\widehat{Y}_{it}^* - \widehat{m}_{i,h}(X_{it})\}^2$.

Proposition S.7. *Let the conditions of Theorem 6.1 be satisfied. Then there exist absolute constants* $0 < c_\nu \le C_\nu < \infty$ *such that*

$$\begin{aligned}
\min_{1\le i\le j\le n}\min_{(x,h)\in\mathcal{G}_T}\sqrt{\widehat{\nu}_{ij}(x,h)} &\ge c_\nu + o_p(1) \\
\max_{1\le i\le j\le n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{\widehat{\nu}_{ij}(x,h)} &\le C_\nu + o_p(1).
\end{aligned}$$

Proof of Proposition S.7. The proposition is a straightforward consequence of the following three observations:

(a) Under our conditions, the term $s(x,h)$ is bounded away from zero and infinity uniformly over (x,h), that is, $0 < c_s \le s(x,h) \le C_s < \infty$ for some absolute constants c_s and C_s.

(b) It holds that

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\big|\widehat{f}_{i,h}(x) - f_i(x)\big| = O_p\Big(\sqrt{\frac{\log n + \log T}{Th_{\min}}} + h_{\max}\Big),$$

where the densities f_i are uniformly bounded away from zero and infinity by (C2).

(c) It holds that

$$\widehat{\sigma}_{i,h}^2 = \sigma_i^2 + b_i^\sigma + R_{i,h}^\sigma \quad \text{with} \quad \max_{1\le i\le n} \max_{\{h:(x,h)\in\mathcal{G}_T\}} |R_{i,h}^\sigma| = o_p(1),$$

where $b_i^\sigma = \mathbb{E}[(\overline{m}_t^{(i)} + \overline{\varepsilon}_t^{(i)})^2]$ and the error variances σ_i^2 are uniformly bounded away from zero and infinity according to (C3). Note that $0 \le b_i^\sigma \le C_b < \infty$ for some sufficiently large absolute constant C_b and that $\max_{1\le i\le n} b_i^\sigma = o(1)$ in the case that n tends to infinity as $T \to \infty$.

Observation (a) can be seen by straightforward arguments and (b) follows from Lemma S.2 together with standard bias calculations. In order to prove (c), we write $\widehat{\sigma}_{i,h}^2 = \sigma_i^2 + b_i^\sigma + R_{i,h}^\sigma$ with $R_{i,h}^\sigma = R_{i,h,1}^\sigma + \ldots + R_{i,h,5}^\sigma$, where

$$\begin{aligned}
R_{i,h,1}^\sigma &= \frac{1}{T}\sum_{t=1}^T \big\{\varepsilon_{it}^2 - \mathbb{E}\big[\varepsilon_{it}^2\big]\big\} \\
R_{i,h,2}^\sigma &= \frac{1}{T}\sum_{t=1}^T \big\{(\overline{m}_t^{(i)} + \overline{\varepsilon}_t^{(i)})^2 - \mathbb{E}\big[(\overline{m}_t^{(i)} + \overline{\varepsilon}_t^{(i)})^2\big]\big\} \\
R_{i,h,3}^\sigma &= \frac{1}{T}\sum_{t=1}^T \big\{\widehat{\Delta}_{i,h}(X_{it}) - (\overline{m}_i + \overline{\varepsilon}_i) + (\overline{\overline{m}}^{(i)} + \overline{\overline{\varepsilon}}^{(i)})\big\}^2 \\
R_{i,h,4}^\sigma &= -\frac{2}{T}\sum_{t=1}^T \big\{\overline{m}_t^{(i)} + \overline{\varepsilon}_t^{(i)}\big\}\big\{\widehat{\Delta}_{i,h}(X_{it}) - (\overline{m}_i + \overline{\varepsilon}_i) + (\overline{\overline{m}}^{(i)} + \overline{\overline{\varepsilon}}^{(i)})\big\} \\
R_{i,h,5}^\sigma &= \frac{2}{T}\sum_{t=1}^T \varepsilon_{it}\big\{\widehat{\Delta}_{i,h}(X_{it}) - (\overline{m}_i + \overline{\varepsilon}_i) - (\overline{m}_t^{(i)} + \overline{\varepsilon}_t^{(i)}) + (\overline{\overline{m}}^{(i)} + \overline{\overline{\varepsilon}}^{(i)})\big\}
\end{aligned}$$

with the shorthand $\widehat{\Delta}_{i,h}(X_{it}) = m_i(X_{it}) - \widehat{m}_{i,h}(X_{it})$. A simplified version of Proposition S.1 yields that

$$\max_{1\le i\le n} \Big| \frac{1}{T}\sum_{t=1}^T \big\{\varepsilon_{it}^2 - \mathbb{E}\big[\varepsilon_{it}^2\big]\big\}\Big| = O_p\Big(\sqrt{\frac{\log n + \log T}{T}}\Big). \tag{S.33}$$

By (C1) and Theorem 5.1(a) in Bradley (2005), the collection of random variables $\mathcal{A}_{i,T} = \{(\varepsilon_{it}, \overline{\varepsilon}_t^{(i)}, \overline{m}_t^{(i)}) : 1 \le t \le T\}$ is strongly mixing for any i and T, where the mixing coefficients $\alpha_{i,T}(\ell)$ of $\mathcal{A}_{i,T}$ are such that $\alpha_{i,T}(\ell) \le n\,\alpha(\ell)$ with $\alpha(\ell)$ decaying to zero exponentially fast. For this reason, we can once again apply a simplified version of Proposition S.1 to obtain that

$$\max_{1\le i\le n} \Big| \frac{1}{T}\sum_{t=1}^T \varepsilon_{it}\big(\overline{m}_t^{(i)} + \overline{\varepsilon}_t^{(i)}\big)\Big| = O_p\Big(\sqrt{\frac{\log n + \log T}{T}}\Big) \tag{S.34}$$

$$\max_{1\le i\le n}\Big|\frac{1}{T}\sum_{t=1}^{T}\Big\{(\overline{m}_t^{(i)}+\overline{\varepsilon}_t^{(i)})^2-\mathbb{E}\big[(\overline{m}_t^{(i)}+\overline{\varepsilon}_t^{(i)})^2\big]\Big\}\Big| = O_p\Big(\sqrt{\frac{\log n+\log T}{T}}\Big). \tag{S.35}$$

Moreover, slightly modifying the proof of Proposition S.6, we can infer that

$$\max_{1\le i\le n}\max_{(x,h)\in\mathcal{G}_T}\big|\widehat{\Delta}_{i,h}(x)\big| = O_p\Big(\sqrt{\frac{\log n+\log T}{Th_{\min}}}+h_{\max}^2\Big). \tag{S.36}$$

Finally, as already seen in the proof of Proposition S.5,

$$\max_{1\le i\le n}\big|\overline{m}_i+\overline{\varepsilon}_i\big| = O_p\Big(\sqrt{\frac{\log n+\log T}{T}}\Big) \tag{S.37}$$

$$\max_{1\le i\le n}\big|\overline{\overline{m}}^{(i)}+\overline{\overline{\varepsilon}}^{(i)}\big| = O_p\Big(\sqrt{\frac{\log n+\log T}{T}}\Big). \tag{S.38}$$

With the help of (S.33)–(S.38), it is not difficult to infer that

$$\max_{1\le i\le n}\max_{\{h:(x,h)\in\mathcal{G}_T\}}|R^{\sigma}_{i,h,\ell}| = o_p(1) \tag{S.39}$$

for $1\le\ell\le 5$, which implies (c). $\square$

S.4 Proof of Theorem 6.1

Proof of (6.2). From Proposition S.5, it follows that

$$\begin{aligned}
&\sqrt{Th}\big\{\widehat{m}_{i,h}(x)-\widehat{m}_{j,h}(x)\big\}\\
&\quad=\sqrt{Th}\big\{m_i(x)-m_j(x)\big\}\\
&\qquad+\sqrt{Th}\Big\{\frac{Q_i^{m,*}(x,h)}{Q_i^*(x,h)}-\frac{Q_j^{m,*}(x,h)}{Q_j^*(x,h)}\Big\}+R_{ij}(x,h),
\end{aligned}$$

where $\max_{1\le i\le j\le n}\max_{(x,h)\in\mathcal{G}_T}|R_{ij}(x,h)| = O_p(\sqrt{\log n+\log T})$. Since $Q_i^{m,*}(x,h) = Q_j^{m,*}(x,h)$ and $Q_i^*(x,h)=Q_j^*(x,h)$ for any two time series i and j in the same group G_k under our conditions, this implies that

$$\max_{1\le k\le K_0}\max_{i,j\in G_k}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\big|\widehat{m}_{i,h}(x)-\widehat{m}_{j,h}(x)\big| = O_p\big(\sqrt{\log n+\log T}\big). \tag{S.40}$$

Moreover, by Proposition S.7,

$$\min_{1\le i\le j\le n}\min_{(x,h)\in\mathcal{G}_T}\sqrt{\widehat{\nu}_{ij}(x,h)}\ge c_\nu+o_p(1),$$

where $c_\nu > 0$ is a sufficiently small absolute constant. As a result, we arrive at

$$\begin{aligned}
\max_{1\le k\le K_0}\max_{i,j\in G_k}\widehat{d}_{ij} &\le \max_{1\le k\le K_0}\max_{i,j\in G_k}\Big\{\max_{(x,h)\in\mathcal{G}_T}|\widehat{\psi}_{ij}(x,h)|\Big\} \\
&\le \frac{\max_{1\le k\le K_0}\max_{i,j\in G_k}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\,\big|\widehat{m}_{i,h}(x)-\widehat{m}_{j,h}(x)\big|}{\min_{1\le i\le j\le n}\min_{(x,h)\in\mathcal{G}_T}\sqrt{\widehat{\nu}_{ij}(x,h)}} \\
&= O_p\big(\sqrt{\log n+\log T}\big),
\end{aligned}$$

which completes the proof. $\square$

Proof of (6.3). By Proposition S.6, it holds that

$$\begin{aligned}
&\sqrt{Th}\big\{\widehat{m}_{i,h}(x)-\widehat{m}_{j,h}(x)\big\} \\
&\quad= \sqrt{Th}\big\{m_i(x)-m_j(x)\big\} \\
&\qquad+ \sqrt{Th^5}\frac{\kappa(x,h)}{2}\big\{m_i''(x)-m_j''(x)\big\} + R_{ij}(x,h),
\end{aligned}$$

where $\max_{1\le i\le j\le n}\max_{(x,h)\in\mathcal{G}_T}|R_{ij}(x,h)| = O_p(\sqrt{\log n+\log T}+\sqrt{Th_{\max}^7})$. With the help of this expansion, we can infer that

$$\begin{aligned}
&\min_{1\le k<k'\le K_0}\min_{\substack{i\in G_k,\\ j\in G_{k'}}}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\,\big|\widehat{m}_{i,h}(x)-\widehat{m}_{j,h}(x)\big| \\
&\quad\ge \min_{1\le k<k'\le K_0}\min_{\substack{i\in G_k,\\ j\in G_{k'}}}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\,\big|m_i(x)-m_j(x)\big| \\
&\qquad- \max_{1\le i\le j\le n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th^5}\,\frac{|\kappa(x,h)|}{2}\,\big|m_i''(x)-m_j''(x)\big| \\
&\qquad- \max_{1\le i\le j\le n}\max_{(x,h)\in\mathcal{G}_T}\big|R_{ij}(x,h)\big| \\
&\quad= \min_{1\le k<k'\le K_0}\min_{\substack{i\in G_k,\\ j\in G_{k'}}}\max_{(x,h)\in\mathcal{G}_T}\sqrt{Th}\,\big|m_i(x)-m_j(x)\big| \\
&\qquad+ O_p\big(\sqrt{Th_{\max}^5}+\sqrt{\log n+\log T}\big) \\
&\quad\ge c\sqrt{Th_{\max}} + o_p\big(\sqrt{Th_{\max}}\big),
\end{aligned}$$

where $c > 0$ is a sufficiently small absolute constant. Moreover, by Proposition S.7,

$$\max_{1\le i\le j\le n}\max_{(x,h)\in\mathcal{G}_T}\sqrt{\widehat{\nu}_{ij}(x,h)} \le C_\nu + o_p(1)$$

with $C_\nu > 0$ being an absolute constant that is chosen sufficiently large. As a consequence, we get that

$$
\begin{aligned}
&\min_{1\le k<k'\le K_0} \min_{\substack{i\in G_k,\\ j\in G_{k'}}} \Big\{ \max_{(x,h)\in\mathcal{G}_T} |\widehat{\psi}_{ij}(x,h)| \Big\} \\
&\ge \frac{\min\limits_{1\le k<k'\le K_0} \min\limits_{i\in G_k, j\in G_{k'}} \max\limits_{(x,h)\in\mathcal{G}_T} \sqrt{Th}\, \big|\widehat{m}_{i,h}(x) - \widehat{m}_{j,h}(x)\big|}{\max\limits_{1\le i\le j\le n} \max\limits_{(x,h)\in\mathcal{G}_T} \sqrt{\widehat{\nu}_{ij}(x,h)}} \\
&\ge c_0\sqrt{Th_{\max}} + o_p\big(\sqrt{Th_{\max}}\big) \qquad \text{(S.41)}
\end{aligned}
$$

with some sufficiently small absolute constant c_0. Since $\lambda(2h_{\min}) = O(\sqrt{\log T})$ by the conditions on the bandwidth $h_{\min}$ in (C8), we finally obtain that

$$
\begin{aligned}
\min_{1\le k<k'\le K_0} \min_{\substack{i\in G_k,\\ j\in G_{k'}}} \widehat{d}_{ij} &\ge \min_{1\le k<k'\le K_0} \min_{\substack{i\in G_k,\\ j\in G_{k'}}} \Big\{ \max_{(x,h)\in\mathcal{G}_T} |\widehat{\psi}_{ij}(x,h)| \Big\} - \lambda(2h_{\min}) \\
&= \min_{1\le k<k'\le K_0} \min_{\substack{i\in G_k,\\ j\in G_{k'}}} \Big\{ \max_{(x,h)\in\mathcal{G}_T} |\widehat{\psi}_{ij}(x,h)| \Big\} + O(\sqrt{\log T}) \\
&\ge c_0\sqrt{Th_{\max}} + o_p\big(\sqrt{Th_{\max}}\big),
\end{aligned}
$$

the last line following from (S.41). $\square$

References

Bosq, D. (1998). *Nonparametric statistics for stochastic processes.* New York, Springer.

Bradley, R. C. (1983). Approximation theorems for strongly mixing random variables. *Michigan Mathematical Journal*, **30** 69–81.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, **2** 107–144.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Annals of Statistics*, **33** 1380–1403.

Hansen, B. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, **24** 726–748.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, **17** 571–599.

Vogt, M. and Linton, O. (2017). Classification of non-parametric regression functions in longitudinal data models. *Journal of the Royal Statistical Society: Series B*, **79** 5–27.